OXFORD

ARTICLE

Personalized Prognostic Prediction Models for Breast Cancer Recurrence and Survival Incorporating Multidimensional Data

Xifeng Wu, Yuanqing Ye, Carlos H. Barcenas, Wong-Ho Chow, Qing H. Meng, Mariana Chavez-MacGregor, Michelle A. T. Hildebrandt, Hua Zhao, Xiangjun Gu, Yang Deng, Elizabeth Wagar, Francisco J. Esteva, Debu Tripathy*, Gabriel N. Hortobagyi*

Affiliations of authors: Departments of Epidemiology (XW, YY, WHC, MATH, HZ, XG, YD), Breast Medical Oncology (CHB, MC, DT, GNH), Laboratory Medicine (QHM, EW), and Health Services Research (MC), The University of Texas MD Anderson Cancer Center, Houston, TX; Perlmutter Cancer Center, New York University Langone Medical Center, New York, NY (FJE)

*Authors contributed equally to this work.
Correspondence to: Xifeng Wu, MD, PhD, Department of Epidemiology, Unit 1340, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030 (e-mail: xwu@mdanderson.org).

Abstract

Background: In this study, we developed integrative, personalized prognostic models for breast cancer recurrence and overall survival (OS) that consider receptor subtypes, epidemiological data, quality of life (QoL), and treatment.
Methods: A total of 15 314 women with stage I to III invasive primary breast cancer treated at The University of Texas MD Anderson Cancer Center between 1997 and 2012 were used to generate prognostic models by Cox regression analysis in a two-stage study. Model performance was assessed by calculating the area under the curve (AUC) and calibration analysis and compared with Nottingham Prognostic Index (NPI) and PREDICT.
Results: Host characteristics were assessed for 10 809 women as the discovery population (median follow-up = 6.09 years, 1144 recurrence and 1627 deaths) and 4505 women as the validation population (median follow-up = 7.95 years, 684 recurrence and 1095 deaths). In addition to the known clinical/pathological variables, the model for recurrence included alcohol consumption while the model for OS included smoking status and physical component summary score. The AUCs for recurrence and OS were 0.813 and 0.810 in the discovery and 0.807 and 0.803 in the validation, respectively, compared with AUCs of 0.761 and 0.753 in discovery and 0.777 and 0.751 in validation for NPI. Our model further showed better calibration compared with PREDICT. We also developed race-specific and receptor subtype–specific models with comparable AUCs. Racial disparity was evident in the distributions of many risk factors and clinical presentation of the disease.
Conclusions: Our integrative prognostic models for breast cancer exhibit high discriminatory accuracy and excellent calibration and are the first to incorporate receptor subtype and epidemiological and QoL data.

Breast cancer is a heterogeneous disease with distinct prognoses. To help establish guidelines for clinical decision-making with regard to personalized management of breast cancer, a number of prognostic models focusing on clinical/pathological and, recently, human epidermal growth factor receptor 2 (HER2) and estrogen receptor (ER) status have been developed since the early 1980s (1–9). Currently, guidelines have been published on a wealth of tumor-based receptor markers (10–16) in

determining prognosis and making decisions regarding adjuvant chemotherapy for early-stage hormone receptor–positive and HER2-negative breast cancer (17). However, validation of these published prognostic models using external patient populations has shown only modest discriminatory accuracy (range = 0.59–0.78) (3–5,18–20). Therefore, a large amount of variability in breast cancer clinical outcomes remains unexplained.

Progesterone receptor (PR) was used in combination with ER and HER2 to define receptor subtypes including HmR-positive (ER+ or PR+, HER2-), HER2-positive (HER2+), and triple-negative (ER-, PR-, HER2-). The inclusion of receptor subtype, patients' demographics, and health-related quality of life (QoL) in the model may improve model performance. These factors were rarely accounted for by the published breast cancer prognostic models. Importantly, racial disparities in breast cancer prognosis are well documented. Despite this, most of the existing models were developed based on white patients. Hence, performance of these models in other racial/ethnic groups is still unclear.

Using a comprehensive database containing breast cancer patients' clinical/pathological, epidemiological, and QoL data carefully compiled for over 20 years at The University of Texas MD Anderson Cancer Center (UTMDACC), we aimed to improve the prediction of breast cancer outcomes by developing models overall, as well as models applicable to patients with non-European ancestry and specific receptor subtype. With the large minority population, we also examined racial/ethnic disparities in terms of risk factors and clinical characteristics of the disease.

Methods

Patient Population and Data Collection

This study included 15 314 women identified from the Breast Medical Oncology Institutional Database at UTMDACC and diagnosed with stage I to III invasive primary breast cancer between 1997 and 2012. To minimize the potential effects of referral bias, models were developed using the 10 809 study participants who met National Comprehensive Cancer Network (NCCN) criteria (21,22). These criteria defined eligible subjects who were age 18 years or older and had newly diagnosed breast cancer, had confirmed histology as defined by the American Joint Committee on Cancer (AJCC), and had received some or all forms of primary breast cancer treatment (surgery, chemotherapy, or hormone therapy). Patients with a history of breast cancer were eligible if they did not have a history of distant metastases and met the aforementioned criteria. Patients with a history of other cancer were eligible if they did not have a history of distant metastases or evidence of relapse and met the aforementioned criteria. The remaining 4505 women were used as the validation population. Epidemiologic and demographic data were collected by self-administered questionnaires at the time of registration and within one year of diagnosis. Patients who drank alcoholic beverages (beer, wine, and liquor) regularly (at least one drink per month) were considered alcohol drinkers. Short Form 12 (SF12) questionnaires were collected at the same time as epidemiologic questionnaires and used to construct QoL variables including physical component summary (PCS) and mental component summary (MCS) scores. Clinical and follow-up data were abstracted from the patient's medical records. Written informed consent was obtained from study participants, and the study was approved by UTMDACC's Institutional Review Board.

Statistical Analysis

The end points were recurrence (defined as the local, regional, or distant metastatic recurrence) and overall survival (OS), and we predicted the likelihood of recurrence or survival within a five-year period. For recurrence, patients who died or who were alive at the last follow-up without evidence of recurrence were censored. Time to recurrence was defined as the time from the date of diagnosis to the date of first documented recurrence, the date of last follow-up, the date of death, or five years after diagnosis, whichever came first. Similarly, OS was defined as the time from the date of diagnosis to the date of death, the date of latest follow-up, or five years after diagnosis, whichever came first. The assumption of proportional hazard was tested by the methods of Therneau and Grambsch. Prognostic factors identified through univariate Cox analysis (Supplementary Tables 1 and 2, available online) were subjected to backward stepwise Cox proportional hazards regression analysis to identify statistically significant prognostic factors (P < .05, Wald test) to be included in the final models. Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated for each factor. Prognostic models for recurrence and OS were generated by initiating the model with clinical/pathological variables, then sequentially incorporating epidemiological variables. Interaction between variables was assessed by adding interaction variable to the Cox model. Significance was checked with Bonferroni adjustment for multiple comparisons (P < .05/number of tests). We assessed model performance by generating receiver operating characteristic (ROC) curves and estimating area under the ROC curves (AUC) for our models and Nottingham Prognostic Index (NPI) (1). We assessed model calibration by generating calibration plot using 100 bootstrap runs to estimate the 95% confidence intervals. We compared the predicted and observed five-year events for our models and PREDICT (4) and used test of proportion for the equality of predicted and observed events. Predicted pure risk of developing the event was estimated using the equation

$F(t, X) = 1 - S(t, X) = 1 - [S_0(t)]^{\exp(\sum_{j=1}^{p} b_j x_j)}$. The differences in distribution of patient characteristics among race/ethnic subgroups were assessed by Pearson Chi-square test. Statistical analyses and modeling were performed using SAS (version 9.2; SAS Institute, Cary, NC), STATA (version 14; Stata Co., College Station, TX), and R software packages. All statistical tests were two-sided unless otherwise indicated, and a P value of less than .05 was considered statistically significant.

Results

Study Population Characteristics

The discovery population (demo+clinical) included a total of 10 809 women (1144 recurrence, 1627 deaths) with a median follow-up of 6.09 years and mean age of 54.5 years (range = 19–98 years). Receptor subtype were available for 9738, self-administered questionnaires (EPI) for 8964 patients, QoL for 5350 patients (4673 before treatment, 677 after treatment), and demo+clinical+EPI+QoL for 5060 patients (501 recurrence, 623 death). The validation population included a total of 4505 women (684 recurrence, 1095 deaths) with a median follow-up of 7.95 years and mean age of 54.7 years (range = 20–91 years). Receptor subtype were available for 3541 patients, EPI for 2348 patients, QoL for 1119 patients (449 before treatment, 670 after treatment), and demo+clinical+EPI+QoL for 1013 patients (159 recurrence, 233 death).

Prognostic Model for Breast Cancer Recurrence

We initiated the model building with baseline demographic and clinical/pathological variables in the entire cohort. Demo+clinical model included age, diagnosis year, race/ethnicity, cancer detection mode, nuclear grade, tumor size, number of positive nodes, lymph or vascular invasion (LVI), histology, and adjuvant hormone therapy (Table 1). Including the receptor subtype resulted in a statistically significantly increased (P = .007) risk of recurrence for patients with triple-negative breast cancer. Alcohol consumption was identified in demo+clinical+EPI model. The recurrence-free survival curve for full model by selected factors is shown in Figure 1, A–C. Good concordance and improvement in AUC was observed for the full model in both discovery and validation data with AUCs of 0.813 and 0.807 (Figure 2, A and B), respectively, although there was little improvement in AUC compared with the clinic model. AUCs for NPI in the the discovery and validation data were 0.761 and 0.777, respectively. We also observed good model calibration for the full model (Supplementary Table 3, available online).

We also generated multivariable risk prognostic models for breast cancer recurrence by receptor subtypes (Table 2) and by race (Table 3). For the subtype-specific models, cancer detection mode, nuclear grade, tumor size, and number of positive nodes were consistently associated in all the subtype-specific models, while age, histology, LVI, and alcohol consumption were in the HmR-positive and triple-negative models, race/ethnicity and diagnosis year were in the HmR and HER2-positive models, and chemotherapy was in the HER2-positive model. For race-specific models, number of positive nodes and adjuvant hormone therapy showed consistent associations, while age, LVI, and receptor subtype were in the white model; histology was in the African American model; cancer detection mode, nuclear grade, and tumor size were in the white and African American models; and diagnosis year and alcohol consumption were in the white and Hispanic models.

Prognostic Model for Breast Cancer OS

As shown in Supplementary Table 4 (available online), the final OS demo+clinical models included terms for age, race/ethnicity, detection mode, nuclear grade, tumor size, number of positive nodes, LVI, receptor subtype, chemotherapy, and adjuvant hormonal therapy. Smoking status and PCS score were identified in the demo+clinical+EPI+QoL model. The OS curve for full model by selected factors is shown in Figure 1, D–F. Excellent model discrimination was observed for the full model in discovery and validation data, with AUCs of 0.810 and 0.803, respectively (Figure 2, C and D), although there was little improvement in AUC compared with the clinic model. The AUCs for NPI in the discovery and validation data were 0.753 and 0.751, respectively. The full model was well calibrated, except for the overall in the validation data (Supplementary Table 3, available online), while PREDICT overestimated the five-year death for the overall and the majority of subgroups in the discovery data and for the triple-negative subgroup in the validation data.

Additional analysis for multivariable risk prognostic models by receptor subtypes and by race was performed (Supplementary Tables 5 and 6, available online). For subtype-specific models, tumor size and number of positive nodes had consistent association, while age, race/ethnicity, and nuclear grade were in the HmR-positive and triple-negative models,

menopausal status and prior cancer were in the HER2-positive model, cancer detection mode and adjuvant hormone therapy were in the HmR and HER2-positive models, LVI was in the triple-negative model, and chemotherapy and PCS score were in the HmR-positive model. For the race-specific model, nuclear grade, tumor size, number of positive nodes, and adjuvant hormone therapy exhibited consistent association, while age, LVI, chemotherapy, and smoking status were in the white model, cancer detection mode and receptor subtype were the in the white and African American models, and PCS score was in the white and Hispanic models.

Application of Prognostic Model for Personalized Risk Prediction

We used the full models to predict pure five-year risk of recurrence and death using hypothetical patients age 40 and 65 years with various risk profiles for HmR-positive and triple-negative breast cancer. For a patient age 40 years with HmR-positive breast cancer and a lowest-risk profile, predicted five-years risk of recurrence was 1.5% (95% CI = 0.9% to 2.0%) for African American, 1.0% (95% CI = 0.6% to 1.3%) for white, 0.9% (95% CI = 0.4% to 1.4%) for Asian/Pacific Islanders, and 0.8% (95% CI = 0.4% to 1.1%) for Hispanic (Figure 3). With the risk profiles changed to diagnosis years 1997–2002, cancer detected by symptoms, tumor size 4–5 cm, nuclear grade III, presence of LVI, five to eight positive nodes, and nonalcohol drinking, predicted five-year risk of recurrence increased to 75.0% (95% CI = 55.6% to 85.9%) for African American, 59.8% (95% CI = 42.0% to 72.1%) for white, 57.7% (95% CI = 33.4% to 73.1%) for Asian/Pacific Islanders, and 51.4% (95% CI = 33.2% to 64.6%) for Hispanic.

For a patient age 40 years with HmR-positive breast cancer, predicted five-year risk of death was 1.2% (95% CI = 0.6% to 1.9%) for African American, 0.8% (95% CI = 0.4% to 1.2%) for white, 0.8% (95% CI = 0.2% to 1.4%) for Asian/Pacific Islanders, and 0.5% (95% CI = 0.2% to 0.8%) for Hispanic with lowest-risk profile (Supplementary Figure 1, available online). When the risk profiles changed to breast cancer detected by symptoms, tumor size 2–3 cm, nuclear grade III, presence of LVI, and three to four positive nodes, predicted five-year risk of death increased to 45.5% (95% CI = 21.9% to 61.9%) for African American, 32.7% (95% CI = 16.4% to 45.7%) for white, 31.7% (95% CI = 6.9% to 49.8%) for Asian/Pacific Islanders, and 20.7% (95% CI = 7.8% to 31.8%) for Hispanic.

Racial/Ethnic Disparities

The highest percentage of recurrence and death was observed for African American women (recurrence = 17.7%; death = 22.3%), followed by white (recurrence = 9.8%; death = 15.4%), Hispanic (recurrence = 10.1%; death = 11.0%), and Asian/Pacific women (recurrence = 8.1%; death = 7.1%). The distributions of the demographic, clinical/pathological, epidemiological, and QoL variables in the discovery population among racial/ethnic subgroups are shown in Table 4. We observed statistically significant racial/ethnic disparities for the majority of these variables (Table 4), except for LVI, although statistically significant interaction of LVI with race/ethnicity was observed for risk of death (P = .04, data not shown). For example, African American women were more likely to be obese (57.3% having BMI ≥ 30 kg/m^2) compared with 13.5% to 36.6% for other races/ethnicities. African American women were more likely (23.6%) to be diagnosed with triple-negative breast cancer

Table 1. Multivariable risk models for breast cancer recurrence and estimated relative risk*

Variable	Demo+clinical HR (95% CI)	Demo+clinical+subtype HR (95% CI)	Demo+clinical+EPI HR (95% CI)
Age, y			
<45	1 (ref)	1 (ref)	1 (ref)
45–54	0.70 (0.58 to 0.83)	0.70 (0.58 to 0.84)	0.67 (0.55 to 0.82)
55–64	0.77 (0.64 to 0.93)	0.76 (0.62 to 0.92)	0.73 (0.59 to 0.91)
≥65	0.87 (0.70 to 1.08)	0.88 (0.70 to 1.11)	0.90 (0.71 to 1.15)
Diagnosis year			
1997–2002	1 (ref)	1 (ref)	1 (ref)
2003–2007	1.02 (0.87 to 1.21)	0.96 (0.80 to 1.14)	0.99 (0.80 to 1.21)
2008–2012	0.69 (0.57 to 0.83)	0.65 (0.54 to 0.80)	0.69 (0.55 to 0.86)
Race/ethnicity			
White	1 (ref)	1 (ref)	1 (ref)
Hispanic	0.94 (0.76 to 1.17)	0.92 (0.73 to 1.15)	0.79 (0.61 to 1.02)
African American	1.72 (1.43 to 2.06)	1.72 (1.42 to 2.07)	1.52 (1.24 to 1.87)
Asian/Pacific	1.08 (0.77 to 1.53)	1.11 (0.78 to 1.59)	0.94 (0.63 to 1.41)
Other	1.52 (0.78 to 2.94)	1.44 (0.74 to 2.80)	1.91 (0.85 to 4.28)
Cancer detection mode			
Screens	1 (ref)	1 (ref)	1 (ref)
Symptom	1.89 (1.58 to 2.26)	1.77 (1.47 to 2.14)	1.96 (1.58 to 2.41)
Nuclear grade			
I or II	1 (ref)	1 (ref)	1 (ref)
III	1.98 (1.65 to 2.38)	2.01 (1.66 to 2.45)	2.11 (1.69 to 2.62)
Histology			
Ductal	1 (ref)	1 (ref)	1 (ref)
Lobular	0.66 (0.47 to 0.92)	0.67 (0.47 to 0.95)	0.69 (0.48 to 1.01)
Mixed ductal/lobular	1.08 (0.79 to 1.47)	1.09 (0.79 to 1.50)	1.16 (0.81 to 1.65)
Other	0.49 (0.29 to 0.84)	0.50 (0.29 to 0.87)	0.49 (0.26 to 0.92)
Tumor size, cm			
0–1	1 (ref)	1 (ref)	1 (ref)
1–2	0.93 (0.75 to 1.15)	0.89 (0.71 to 1.11)	0.87 (0.68 to 1.10)
2–3	1.18 (0.94 to 1.49)	1.16 (0.91 to 1.47)	1.18 (0.91 to 1.52)
3–4	1.54 (1.18 to 2.00)	1.48 (1.13 to 1.94)	1.31 (0.97 to 1.77)
4–5	1.90 (1.40 to 2.57)	1.91 (1.40 to 2.60)	1.76 (1.25 to 2.46)
5+	2.52 (1.96 to 3.24)	2.42 (1.87 to 3.14)	2.35 (1.78 to 3.10)
No. of positive nodes			
0	1 (ref)	1 (ref)	1 (ref)
1 or 2	1.51 (1.23 to 1.86)	1.52 (1.23 to 1.89)	1.51 (1.20 to 1.91)
3 or 4	2.67 (2.10 to 3.39)	2.65 (2.07 to 3.40)	2.34 (1.78 to 3.08)
5–8	3.96 (3.11 to 5.05)	4.08 (3.17 to 5.25)	4.15 (3.16 to 5.45)
9–19	5.19 (4.08 to 6.60)	5.48 (4.25 to 7.05)	5.44 (4.13 to 7.16)
≥20	7.12 (4.93 to 10.29)	7.25 (5.00 to 10.51)	7.99 (5.39 to 11.83)
Lymph or vascular invasion			
No	1 (ref)	1 (ref)	1 (ref)
Yes	1.50 (1.28 to 1.75)	1.51 (1.29 to 1.78)	1.63 (1.36 to 1.94)
Adjuvant hormone therapy			
No	1 (ref)	1 (ref)	1 (ref)
Yes	0.40 (0.34 to 0.46)	0.46 (0.37 to 0.56)	0.48 (0.38 to 0.61)
Breast cancer receptor subtype			
HmR-positive	—	1 (ref)	1 (ref)
HER2-positive	—	0.89 (0.71 to 1.11)	0.88 (0.69 to 1.12)
Triple-negative	—	1.40 (1.10 to 1.79)	1.58 (1.20 to 2.07)
Alcohol consumption			
No	—	—	1 (ref)
Yes	—	—	0.75 (0.64 to 0.88)
Discovery			
No. of events/total population	817/10 013	759/9035	642/7760
AUC	0.805	0.807	0.813
Validation			
No. of events/total population	335/3743	267/2981	151/1717
AUC	0.798	0.812	0.807

*AUC = area under the curve; CI = confidence interval; EPI = self-administered questionnaires; HmR = hormone receptor; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.
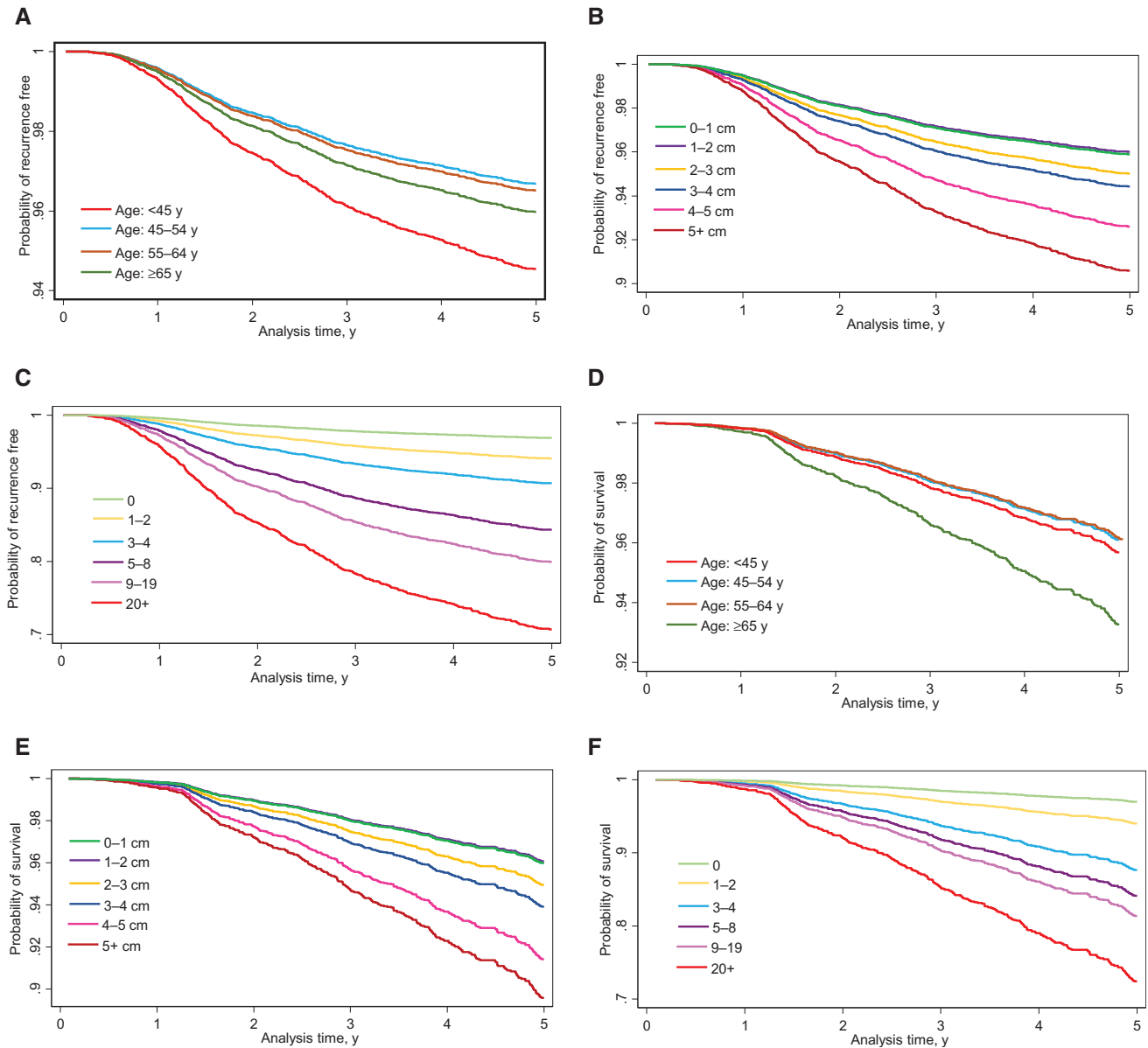
Figure 1. Survival curves based on the full model. Graphs show relapse-free survival based on the full model by **(A)** age; **(B)** tumor size; and **(C)** number of positive nodes. Graphs show overall survival curve based on the full model by **(D)** age; **(E)** tumor size; and **(F)** number of positive nodes.

than other races/ethnicities (range = 11.1% to 12.5%). In terms of QoL, 19.1% of African American women exhibited low PCS, followed by Hispanic (17.2%), white (12.6%), and Asian/Pacific (8.7%). For risk of recurrence, a statistically significant interaction of race/ethnicity with diagnosis year was also observed ($P = .02$).

Discussion

This is the first large-scale study to incorporate epidemiological characteristics and QoL data to the breast cancer prognosis model. Our models, which were developed using the large patient population at UTMDACC, had AUCs of 0.813 for recurrence and 0.810 for survival in the discovery population. The models' high discriminatory accuracy was further confirmed by AUCs of 0.807 for recurrence and 0.803 for survival in the validation population. Calibration analysis demonstrated the good

agreement between the predicted and observed five-year events in both the discovery and validation populations. Further, our models have improved AUCs compared with NPI and better calibrations compared with PREDICT. Using the large patient population, we published for the first time separate models for racial/ethnic subgroups and found racial/ethnic disparities across majorities of demographic, clinical/pathological, epidemiological, and QoL variables.

To date, considerable effort has gone into the development of prognostic models for breast cancer clinical outcomes, particularly to individualize treatment recommendations. Earlier analyses focused on clinical and pathological factors while recent studies included individual receptor status in an effort to more precisely predict of patients' outcome, especially in the context of specific therapies. Nevertheless, no prognostic model for breast cancer outcomes currently uses the receptor subtype information. Our study showed that triple-negative breast
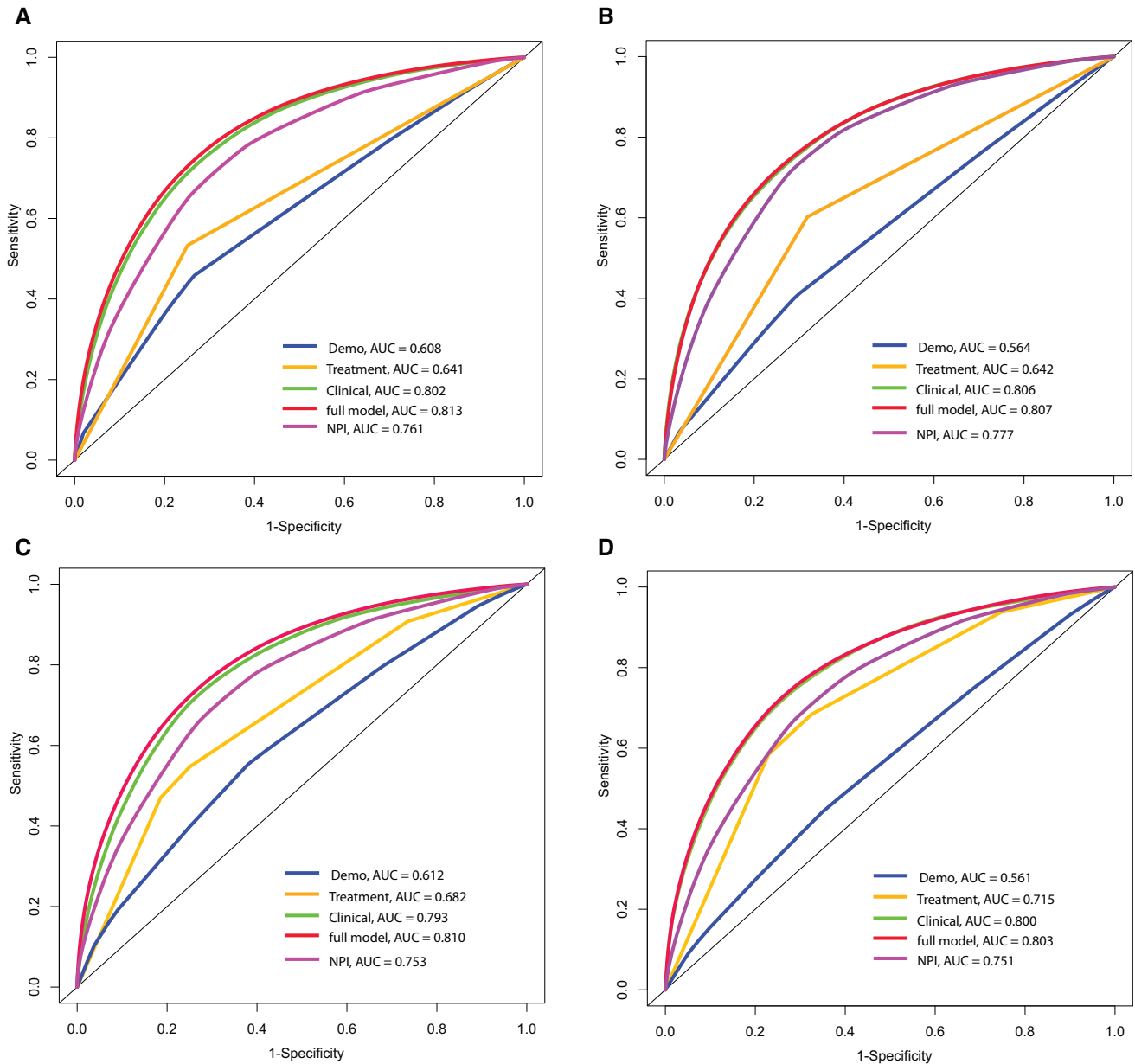
Figure 2. Discriminatory accuracy for predicting breast cancer outcomes assessed by receiver operator characteristics analysis calculating area under the curve. **A)** Recurrence within five years in the discovery population. **B)** Recurrence within five years in the validation population. **C)** Survival within five years in the discovery population. **D)** Survival within five years in the validation population. For recurrence: the demographic model included age and race; clinical model included diagnosis year; nuclear grade, tumor size, histology, number of positive nodes, cancer detection mode, breast cancer subtype, and lymphatic/vascular invasion; treatment model included adjuvant hormone therapy; full model included alcohol consumption in addition to demo, clinical, and treatment variables mentioned above. For survival: the demo model included age and race; clinical model included nuclear grade, tumor size, number of positive nodes, cancer detection mode, breast cancer subtype, and lymph or vascular invasion; treatment model included chemotherapy and adjuvant hormone therapy; full model included smoking status and physical component summary score in addition to the demo, clinical, and treatment variables mentioned above. AUC = area under the curve; NPI = Nottingham Prognostic Index.

cancer patients have a statistically significantly increased risk of recurrence and death, while HER2-positive breast cancer patients have improved OS when compared with HmR-positive breast cancer patients.

Many studies have shown racial/ethnic difference in breast cancer recurrence and survival (24–26). The racial/ethnic disparities in clinical/pathological variables, treatment, family history of breast cancer, and QoL have also been studied (27–30), but few studies have examined the racial/ethnic disparities associated with smoking, alcohol consumption, and personal history of cancer. In this study, we demonstrated racial/ethnic disparities in almost all variables and incorporated race/ethnicity into breast

cancer prognostic models. It is also noteworthy that, in comparison with white women, Hispanic women exhibited a consistently and statistically significantly reduced risk of death, while African American women exhibited a statistically significantly increased risk of recurrence and death. Therefore, the difference between the prognosis of African American and white patients could not be explained by the clinical/pathological differences. Further examination for the interaction of race/ethnicity with other factors found statistically significant interactions with presence of lymph or vascular invasion for risk of death and diagnosis year for risk of recurrence. However, they did not remain statistically significant after Bonferroni correction for multiple comparisons.

Table 2. Multivariable risk models for breast cancer recurrence by breast cancer receptor subtypes*

Variable	Demo+clinical			Demo+clinical+EPI	
	HmR-positive HR (95% CI)	HER2-positive HR (95% CI)	Triple-negative HR (95% CI)	HmR-positive HR (95% CI)	Triple-negative HR (95% CI)
Age, y					
<45	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)
45–54	0.70 (0.53 to 0.92)	—	0.61 (0.44 to 0.85)	0.68 (0.50 to 0.92)	0.58 (0.41 to 0.82)
55–64	0.76 (0.57 to 1.00)	—	0.63 (0.44 to 0.91)	0.75 (0.55 to 1.02)	0.61 (0.41 to 0.90)
≥65	0.68 (0.49 to 0.95)	—	1.17 (0.78 to 1.75)	0.69 (0.48 to 1.00)	1.16 (0.77 to 1.76)
Diagnosis year					
1997–2002	1 (ref)	1 (ref)	—	1 (ref)	—
2003–2007	1.00 (0.77 to 1.30)	0.71 (0.50 to 1.01)	—	1.01 (0.74 to 1.39)	—
2008–2012	0.65 (0.49 to 0.88)	0.44 (0.28 to 0.67)	—	0.68 (0.48 to 0.96)	—
Race/ethnicity					
White	1 (ref)	1 (ref)	—	1 (ref)	—
Hispanic	0.91 (0.66 to 1.26)	0.86 (0.54 to 1.37)	—	0.79 (0.55 to 1.14)	—
African American	1.91 (1.44 to 2.53)	1.72 (1.12 to 2.63)	—	1.53 (1.11 to 2.11)	—
Asian/Pacific	1.12 (0.69 to 1.82)	1.26 (0.61 to 2.61)	—	0.95 (0.56 to 1.61)	—
Other	1.99 (0.74 to 5.37)	0.80 (0.11 to 5.82)	—	1.11 (0.27 to 4.48)	—
Cancer detection mode					
Screens	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Symptom	2.05 (1.56 to 2.68)	1.76 (1.21 to 2.58)	1.48 (1.04 to 2.11)	2.25 (1.66 to 3.04)	1.55 (1.05 to 2.28)
Nuclear grade					
I or II	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
III	2.03 (1.62 to 2.53)	1.92 (1.15 to 3.22)	1.95 (1.04 to 3.67)	2.16 (1.69 to 2.76)	2.14 (1.05 to 4.40)
Histology					
Ductal	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)
Lobular	0.59 (0.40 to 0.88)	—	1.06 (0.25 to 4.44)	0.64 (0.42 to 0.97)	0.47 (0.06 to 3.47)
Mixed ductal/lobular	0.82 (0.56 to 1.22)	—	4.54 (2.01 to 10.25)	0.80 (0.51 to 1.25)	4.76 (2.10 to 10.79)
Other	0.45 (0.21 to 0.96)	—	0.54 (0.17 to 1.70)	0.42 (0.17 to 1.02)	0.37 (0.09 to 1.51)
Tumor size, cm					
0–1	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
1–2	0.95 (0.69 to 1.32)	1.20 (0.76 to 1.90)	0.76 (0.50 to 1.14)	0.91 (0.63 to 1.30)	0.76 (0.49 to 1.17)
2–3	1.17 (0.82 to 1.68)	1.23 (0.73 to 2.08)	1.33 (0.88 to 2.00)	1.22 (0.82 to 1.79)	1.25 (0.81 to 1.94)
3–4	1.74 (1.15 to 2.65)	1.84 (1.04 to 3.27)	1.35 (0.84 to 2.18)	1.46 (0.90 to 2.35)	1.23 (0.75 to 2.03)
4–5	2.46 (1.57 to 3.86)	1.65 (0.77 to 3.56)	1.90 (1.11 to 3.23)	2.27 (1.37 to 3.76)	1.88 (1.08 to 3.28)
5+	2.39 (1.60 to 3.56)	2.55 (1.47 to 4.40)	3.10 (1.98 to 4.84)	2.21 (1.44 to 3.40)	3.08 (1.93 to 4.89)
No. of positive nodes					
0	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
1 or 2	1.65 (1.19 to 2.28)	1.37 (0.82 to 2.27)	1.51 (1.04 to 2.17)	1.70 (1.19 to 2.44)	1.39 (0.95 to 2.05)
3 or 4	2.45 (1.67 to 3.58)	3.58 (2.14 to 5.97)	2.17 (1.37 to 3.42)	2.35 (1.54 to 3.61)	1.79 (1.11 to 2.91)
5–8	4.42 (3.08 to 6.32)	3.50 (2.02 to 6.06)	3.85 (2.41 to 6.14)	4.83 (3.25 to 7.17)	4.06 (2.49 to 6.64)
9–19	5.74 (3.98 to 8.26)	4.61 (2.69 to 7.90)	6.75 (4.25 to 10.71)	6.26 (4.19 to 9.35)	5.93 (3.59 to 9.79)
≥20	8.61 (5.12 to 14.48)	4.02 (1.38 to 11.68)	6.45 (3.40 to 12.23)	9.34 (5.26 to 16.57)	7.67 (3.96 to 14.85)
Lymph or vascular invasion					
No	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)
Yes	1.42 (1.13 to 1.78)	—	1.97 (1.46 to 2.65)	1.47 (1.14 to 1.89)	2.14 (1.56 to 2.93)
Adjuvant hormone therapy					
No	1 (ref)	1 (ref)	—	1 (ref)	—
Yes	0.38 (0.29 to 0.49)	0.53 (0.38 to 0.73)	—	0.39 (0.29 to 0.53)	—
Chemotherapy					
No	—	1 (ref)	—	—	—
Yes	—	0.55 (0.31 to 0.98)	—	—	
Alcohol consumption					
No	—	—	—	1 (ref)	1 (ref)
Yes	—	—	—	0.73 (0.58 to 0.93)	0.73 (0.55 to 0.97)
Discovery					
No. of events/total population	360/6469	163/1384	236/1182	296/5557	213/1025
AUC	0.808	0.752	0.725	0.815	0.719
Validation					
No. of events/total population	107/2043	62/466	98/472	60/1177	60/288
AUC	0.795	0.788	0.715	0.789	0.695

*AUC = area under the curve; CI = confidence interval; EPI = self-administered questionnaires; HmR = hormone receptor; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.

ARTICLE

ARTICLE

Table 3. Multivariable risk models for breast cancer recurrence by race*

| Variable | Demo+clinical | | | Demo+clinical+subtype | Demo+clinical+EPI | |
	White HR (95% CI)	Hispanic HR (95% CI)	African American HR (95% CI)	White HR (95% CI)	White HR (95% CI)	Hispanic HR (95% CI)
Age, y						
<45	1 (ref)	—	—	1 (ref)	1 (ref)	1 (ref)
45–54	0.70 (0.55 to 0.88)	—	—	0.68 (0.53 to 0.87)	0.66 (0.51 to 0.86)	—
55–64	0.78 (0.62 to 0.98)	—	—	0.76 (0.59 to 0.97)	0.75 (0.58 to 0.98)	—
≥65	0.82 (0.63 to 1.07)	—	—	0.86 (0.66 to 1.13)	0.88 (0.66 to 1.18)	—
Diagnosis year						
1997–2002	1 (ref)	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)
2003–2007	1.00 (0.82 to 1.22)	0.77 (0.48 to 1.24)	—	0.95 (0.76 to 1.17)	0.84 (0.66 to 1.08)	0.87 (0.49 to 1.54)
2008–2012	0.68 (0.54 to 0.85)	0.43 (0.25 to 0.72)	—	0.65 (0.51 to 0.83)	0.60 (0.45 to 0.79)	0.37 (0.19 to 0.71)
Cancer detection mode						
Screens	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)	—
Symptom	2.06 (1.64 to 2.60)	—	1.76 (1.19 to 2.61)	1.96 (1.54 to 2.49)	2.11 (1.61 to 2.75)	—
Nuclear grade						
I or II	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)	—
III	2.25 (1.81 to 2.80)	—	1.63 (1.02 to 2.59)	2.27 (1.78 to 2.90)	2.20 (1.69 to 2.87)	—
Histology						
Ductal			1 (ref)			
Lobular	—	—	0.12 (0.03 to 0.51)	—	—	—
Mixed ductal/lobular	—	—	1.31 (0.57 to 2.99)	—	—	—
Other	—	—	0.45 (0.14 to 1.43)	—	—	—
Tumor size, cm						
0–1	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)	—
1–2	0.96 (0.73 to 1.25)	—	0.77 (0.45 to 1.30)	0.90 (0.68 to 1.19)	0.92 (0.68 to 1.24)	—
2–3	1.29 (0.97 to 1.72)	—	1.20 (0.71 to 2.03)	1.25 (0.93 to 1.68)	1.34 (0.97 to 1.84)	—
3–4	1.52 (1.09 to 2.12)	—	2.25 (1.30 to 3.92)	1.44 (1.02 to 2.04)	1.29 (0.88 to 1.88)	—
4–5	2.25 (1.54 to 3.27)	—	2.32 (1.19 to 4.51)	2.23 (1.52 to 3.28)	2.13 (1.39 to 3.28)	—
5+	2.24 (1.62 to 3.11)	—	4.31 (2.53 to 7.35)	2.15 (1.53 to 3.01)	2.09 (1.46 to 3.00)	—
No. of positive nodes						
0	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
1 or 2	1.47 (1.14 to 1.90)	1.74 (0.91 to 3.35)	1.51 (0.94 to 2.42)	1.48 (1.13 to 1.93)	1.48 (1.11 to 1.98)	1.48 (0.71 to 3.10)
3 or 4	2.29 (1.69 to 3.10)	3.56 (1.74 to 7.25)	3.73 (2.25 to 6.19)	2.29 (1.67 to 3.15)	1.89 (1.33 to 2.69)	2.85 (1.17 to 6.91)
5–8	3.64 (2.69 to 4.93)	8.40 (4.31 to 16.37)	3.68 (2.08 to 6.51)	3.70 (2.70 to 5.06)	3.82 (2.73 to 5.33)	7.91 (3.55 to 17.59)
9–19	4.21 (3.11 to 5.71)	12.31 (6.44 to 23.54)	6.69 (4.04 to 11.07)	4.52 (3.27 to 6.23)	4.41 (3.10 to 6.26)	16.50 (7.99 to 34.09)
≥20	5.99 (3.78 to 9.51)	15.82 (6.63 to 37.76)	16.11 (6.80 to 38.16)	5.87 (3.68 to 9.36)	6.81 (4.19 to 11.05)	19.43 (7.80 to 48.39)
Lymph or vascular invasion						
No	1 (ref)	1 (ref)	—	1 (ref)	1 (ref)	1 (ref)
Yes	1.51 (1.24 to 1.84)	1.58 (1.02 to 2.47)	—	1.57 (1.28 to 1.93)	1.81 (1.45 to 2.26)	1.32 (0.79 to 2.22)
Adjuvant hormone therapy						
No	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Yes	0.39 (0.33 to 0.47)	0.23 (0.16 to 0.35)	0.45 (0.32 to 0.65)	0.47 (0.36 to 0.60)	0.49 (0.36 to 0.65)	0.20 (0.12 to 0.32)

(continued)

Table 3. (continued)

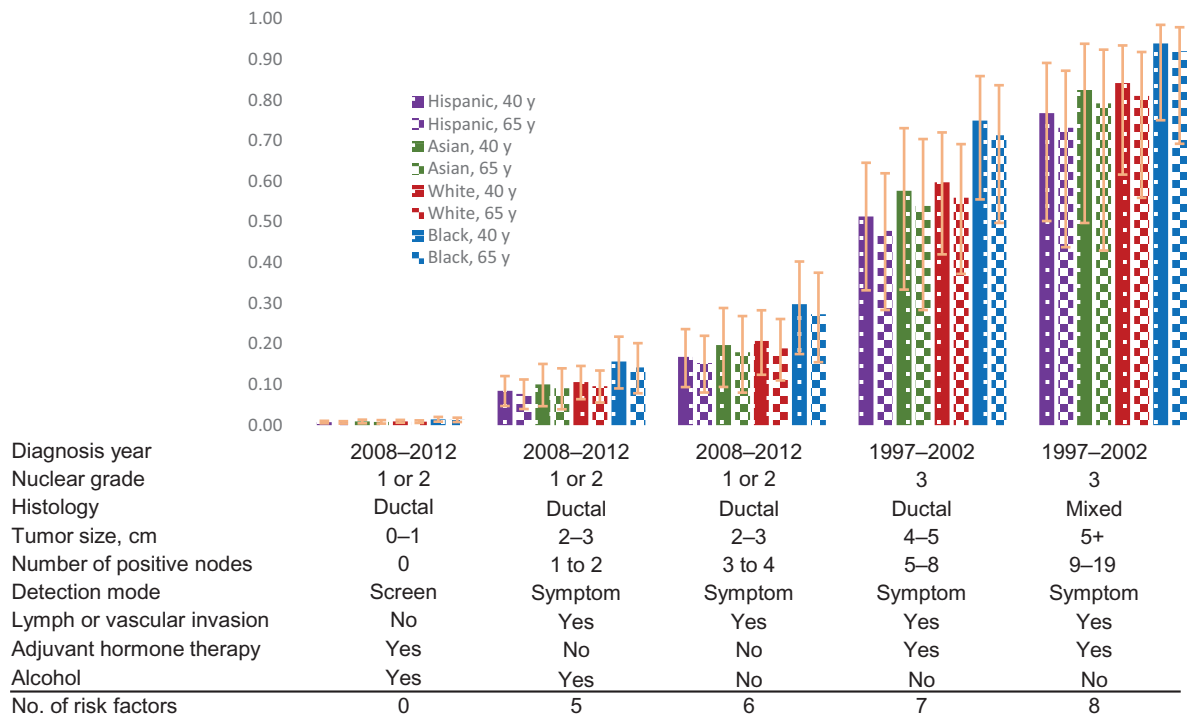|	Demo+clinical			Demo+clinical+subtype	Demo+clinical+EPI	
Variable	White HR (95% CI)	Hispanic HR (95% CI)	African American HR (95% CI)	White HR (95% CI)	White HR (95% CI)	Hispanic HR (95% CI)
Breast cancer receptor subtype						
HmR-positive	—	—	—	1 (ref)	1 (ref)	—
HER2-positive	—	—	—	0.93 (0.71 to 1.23)	0.91 (0.67 to 1.24)	—
Triple-negative	—	—	—	1.46 (1.07 to 2.00)	1.64 (1.16 to 2.31)	—
Alcohol consumption						
No	—	—	—	—	1 (ref)	1 (ref)
Yes	—	—	—	—	0.77 (0.63 to 0.94)	0.52 (0.31 to 0.87)
Discovery						
No. of events/total population	520/7152	99/1278	154/1003	478/6419	410/5456	77/1080
AUC	0.802	0.790	0.793	0.805	0.809	0.802
Validation						
No. of events/total population	238/2524	37/519	51/516	185/1997	108/1160	19/253
AUC	0.803	0.838	0.748	0.818	0.811	0.766

*AUC = area under the curve; CI = confidence interval; EPI = self-administered questionnaires; HmR = hormone receptor; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.

Further, compared with existing models, our models incorporated many more additional epidemiological variables as independent prognostic factors including race/ethnicity, alcohol consumption, and smoking status. Prior meta-analyses have reported statistically significant associations between alcohol consumption and a reduced risk of death (31) and between smoking and an increased risk of death (32). However, the association between alcohol consumption and the risk of recurrence has previously been unclear (31,33,34). Our study suggested that alcohol consumption protects against recurrence. In addition, the association between smoking and a greater risk of death was confirmed by our study.

Health-related QoL has a profound impact on an individual's health. A statistically significant association between physical functioning and survival has been reported by several studies (35–39), while other studies have found no statistically significant association (40–42). The discrepancy in these findings may be attributed to the differences in QoL instruments, the timing of QoL assessments, racial/ethnic background, breast cancer patient characteristics, and the type of study (single- or multi-institution), etc. In this study, we found that a low PCS score was associated with a statistically significantly increased risk of death. Therefore, our study provided evidence for the effect of physical well-being and illustrated the disparity in QoL among breast cancer patients.

Our models were developed with a large patient population with discovery and validation phases. Other more overt biases exist such as the use of adjuvant therapy (particularly chemotherapy) in higher-risk patients. Therefore, external validation in independent and diverse patient populations is also needed before our models can be applied in the clinical setting. Further, incorporation of epidemiological and QoL variables into our final models appears to marginally improve the models' discriminatory accuracy compared with clinic models, consistent with the observation from another study (43). However, calibration analysis showed that full models were well calibrated for almost all the subgroup analysis except for two scenarios, while the clinic model did not have good calibration for 10 scenarios. Decision curve analysis further demonstrated higher net benefit for the full model compared with the clinical model. Because UTMDACC is a tertiary cancer center, patients who are not initially treated at UTMDACC and have poor prognostic features, such as tumor progression during neoadjuvant chemotherapy, among many other reasons, will further seek treatment at UTMDACC, as opposed to those who do not have these poor prognostic features. Therefore, there is a potential for referral bias. We focused the discovery data on patients who met NCCN criteria in order to broaden the applicability of the study in a hospital setting and used the remaining patients as validation data to examine whether the models are applicable to a different patient population with more aggressive disease. However, because only data from a single institution was used, it may limit the potential generalizability to external populations, and further validations in external populations are needed in order to generalize our findings. The median age of our patient population was about 10 years younger than that of overall US breast cancer population, and our populations also had a higher-than-average number of triple-negative and HER2-positive breast cancers. Also, the treatment plan might change during the study period, which could influence the patient prognosis, and in this retrospective analysis we did not have access to the reasons why patients received their specific treatments, although treatment of patients from single institution might be more homogeneous. We combined the locoregional and distant
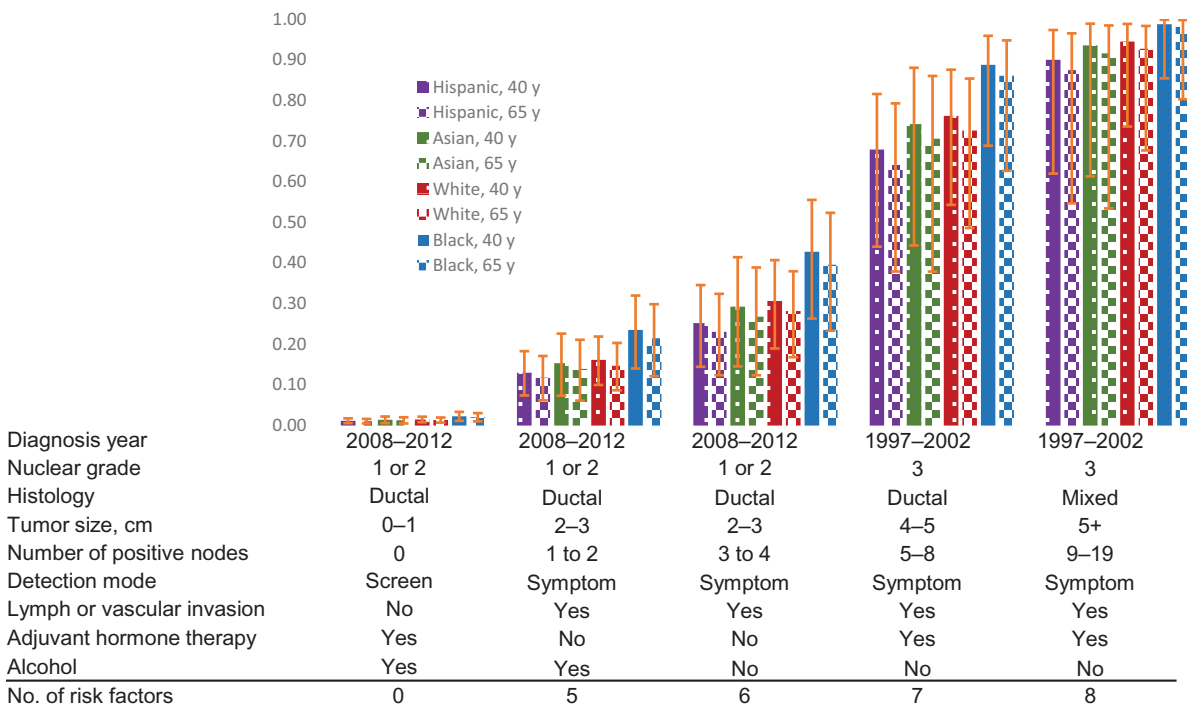
ARTICLE

A

Diagnosis year	2008–2012	2008–2012	2008–2012	1997–2002	1997–2002
Nuclear grade	1 or 2	1 or 2	1 or 2	3	3
Histology	Ductal	Ductal	Ductal	Ductal	Mixed
Tumor size, cm	0–1	2–3	2–3	4–5	5+
Number of positive nodes	0	1 to 2	3 to 4	5–8	9–19
Detection mode	Screen	Symptom	Symptom	Symptom	Symptom
Lymph or vascular invasion	No	Yes	Yes	Yes	Yes
Adjuvant hormone therapy	Yes	No	No	Yes	Yes
Alcohol	Yes	Yes	No	No	No
No. of risk factors	0	5	6	7	8

B

Diagnosis year	2008–2012	2008–2012	2008–2012	1997–2002	1997–2002
Nuclear grade	1 or 2	1 or 2	1 or 2	3	3
Histology	Ductal	Ductal	Ductal	Ductal	Mixed
Tumor size, cm	0–1	2–3	2–3	4–5	5+
Number of positive nodes	0	1 to 2	3 to 4	5–8	9–19
Detection mode	Screen	Symptom	Symptom	Symptom	Symptom
Lymph or vascular invasion	No	Yes	Yes	Yes	Yes
Adjuvant hormone therapy	Yes	No	No	Yes	Yes
Alcohol	Yes	Yes	No	No	No
No. of risk factors	0	5	6	7	8

Figure 3. Application of the prediction models to predict pure risk of developing breast cancer recurrence in five years for hypothetical individuals with different risk profiles. **A)** predicted five-year risk of recurrence for HmR-positive (estrogen receptor–positive, progesterone receptor–positive, human epidermal growth factor receptor–negative) female breast cancer patients age 40 and 65 years. **B)** Predicted five-year risk of recurrence for triple-negative female breast cancer patients age 40 and 65 years. **Error bars** represent 95% confidence interval.

ARTICLE

Table 4. Racial/ethnic disparities of demographic, clinical, epidemiological, and QoL risk factors

Variable	White No. (%)	African American No. (%)	Hispanic No. (%)	Asian/ Pacific No. (%)	P*
Age, y					
<45	1453 (0.19)	261 (0.24)	400 (0.29)	155 (0.3)	<.001
45–54	2262 (0.29)	371 (0.34)	455 (0.33)	173 (0.33)	
55–64	2152 (0.28)	247 (0.23)	349 (0.25)	113 (0.22)	
≥65	1853 (0.24)	213 (0.2)	172 (0.13)	78 (0.15)	
BMI, kg/m^2					
<25	2946 (0.39)	151 (0.14)	395 (0.29)	283 (0.55)	<.001
25–29.9	2335 (0.31)	302 (0.28)	467 (0.34)	159 (0.31)	
≥30	2281 (0.3)	610 (0.57)	497 (0.37)	69 (0.14)	
Menopausal status					
Pre	2306 (0.3)	345 (0.32)	569 (0.41)	250 (0.48)	<.001
Peri	228 (0.03)	26 (0.02)	35 (0.03)	12 (0.02)	
Post	5179 (0.67)	721 (0.66)	771 (0.56)	257 (0.5)	
Cancer detection mode					
Screens	3864 (0.5)	479 (0.44)	549 (0.4)	224 (0.43)	<.001
Symptom	3845 (0.5)	613 (0.56)	825 (0.6)	295 (0.57)	
Stage					
I	3864 (0.5)	453 (0.41)	582 (0.42)	239 (0.46)	<.001
II	2933 (0.38)	461 (0.42)	576 (0.42)	213 (0.41)	
III	923 (0.12)	178 (0.16)	218 (0.16)	67 (0.13)	
Nuclear grade					
I or II	4390 (0.58)	385 (0.36)	703 (0.52)	287 (0.55)	<.001
III	3223 (0.42)	696 (0.64)	656 (0.48)	231 (0.45)	
Histology					
Ductal	6039 (0.78)	932 (0.85)	1113 (0.81)	424 (0.82)	<.001
Lobular	729 (0.09)	79 (0.07)	113 (0.08)	38 (0.07)	
Mixed ductal/lobular	569 (0.07)	37 (0.03)	90 (0.07)	28 (0.05)	
Other	383 (0.05)	44 (0.04)	60 (0.04)	29 (0.06)	
Tumor size, cm					
0–1	2219 (0.3)	299 (0.29)	323 (0.25)	158 (0.32)	<.001
1–2	2925 (0.4)	345 (0.34)	487 (0.38)	159 (0.32)	
2–3	1129 (0.15)	185 (0.18)	252 (0.19)	89 (0.18)	
3–4	452 (0.06)	82 (0.08)	94 (0.07)	36 (0.07)	
4–5	211 (0.03)	36 (0.04)	45 (0.03)	18 (0.04)	
5+	404 (0.06)	73 (0.07)	96 (0.07)	30 (0.06)	
No. of positive nodes					
0	4501 (0.59)	569 (0.52)	741 (0.54)	290 (0.56)	<.001
1 or 2	1700 (0.22)	259 (0.24)	314 (0.23)	127 (0.25)	
3 or 4	573 (0.08)	97 (0.09)	129 (0.09)	40 (0.08)	
5–8	424 (0.06)	72 (0.07)	91 (0.07)	35 (0.07)	
9–19	349 (0.05)	76 (0.07)	75 (0.05)	21 (0.04)	
≥20	92 (0.01)	13 (0.01)	22 (0.02)	3 (0.01)	
Lymph or vascular invasion					
No	5959 (0.77)	831 (0.76)	1027 (0.75)	392 (0.76)	.14
Yes	1739 (0.23)	258 (0.24)	348 (0.25)	126 (0.24)	
Breast cancer receptor subtype					
HR-positive	5100 (0.74)	594 (0.6)	864 (0.68)	340 (0.71)	<.001
HER2-positive	984 (0.14)	157 (0.16)	253 (0.2)	86 (0.18)	
Triple-negative	824 (0.12)	232 (0.24)	160 (0.13)	53 (0.11)	
Chemotherapy					
No	3062 (0.4)	330 (0.3)	393 (0.29)	180 (0.35)	<.001
Yes	4658 (0.6)	762 (0.7)	983 (0.71)	339 (0.65)	
Adjuvant hormone therapy					
No	2125 (0.28)	449 (0.41)	389 (0.28)	139 (0.27)	<.001
Yes	5595 (0.72)	643 (0.59)	987 (0.72)	380 (0.73)	
Radiotherapy					
No	2685 (0.35)	336 (0.31)	488 (0.35)	187 (0.36)	.04
Yes	5035 (0.65)	756 (0.69)	888 (0.65)	332 (0.64)	
Personal cancer history					
No	5482 (0.91)	891 (0.98)	1120 (0.97)	447 (0.98)	<.001
Yes	550 (0.09)	18 (0.02)	31 (0.03)	8 (0.02)	

(continued)

ARTICLE

Table 4. (continued)

Variable	White No. (%)	African American No. (%)	Hispanic No. (%)	Asian/ Pacific No. (%)	P*
Smoking status					
Never	3639 (0.59)	646 (0.69)	842 (0.72)	426 (0.92)	<.001
Former	1965 (0.32)	202 (0.21)	236 (0.2)	27 (0.06)	
Current	590 (0.1)	93 (0.1)	84 (0.07)	12 (0.03)	
Alcohol consumption					
No	2732 (0.44)	580 (0.62)	718 (0.62)	376 (0.82)	<.001
Yes	3447 (0.56)	353 (0.38)	444 (0.38)	84 (0.18)	
Family history of any cancer*					
No	2716 (0.43)	466 (0.49)	669 (0.56)	297 (0.63)	<.001
Yes	3566 (0.57)	485 (0.51)	516 (0.44)	175 (0.37)	
Family history of breast cancer*					
No	5161 (0.82)	773 (0.81)	1006 (0.85)	414 (0.88)	.002
Yes	1121 (0.18)	178 (0.19)	179 (0.15)	58 (0.12)	
PCS score					
High	1895 (0.50)	168 (0.31)	275 (0.42)	143 (0.50)	<.001
Medium	1450 (0.38)	262 (0.50)	267 (0.41)	120 (0.42)	
Low	483(0.13)	104 (0.19)	113 (0.17)	25 (0.09)	
MCS score					
High	290 (0.08)	68 (0.13)	56 (0.09)	34 (0.12)	.001
Medium	2611 (0.68)	333 (0.62)	415 (0.63)	190 (0.66)	
Low	927 (0.24)	133 (0.25)	184 (0.28)	64 (0.22)	

*Chi-square test (two-sided). BMI = body mass index; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; MCS = mental component summary; PCS = physical component summary; QoL = quality of life.

recurrence as overall recurrence because of the small number of events for locoregional recurrence. The results for distant recurrence were similar to overall recurrence in the univariate analysis (data not shown).

In conclusion, we have developed and validated prognostic models for early-stage breast cancer patients by adding, for the first time, receptor subtype and epidemiological and QoL variables. Our models have high discriminatory accuracy and excellent calibration. We have also demonstrated the health and well-being disparities between breast cancer patients of different racial/ethnic identifies.

Funding

Notes

The study sponsor had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

References

1. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast-cancer. *Br J Cancer*. 1982;45(3):361–366.
2. Ravdin PM, Siminoff LA, Davis GJ, et al.Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19(4):980–991.
3. Wishart GC, Azzato EM, Greenberg DC, et al.PREDICT: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12(1):R1.
4. Wishart GC, Bajdik CD, Dicks E, et al.PREDICT Plus: Development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer*. 2012;107(5):800–807.
5. Kim W, Kim KS, Lee JE, et al.Development of novel breast cancer recurrence prediction model using support vector machine. *J Breast Cancer*. 2012;15(2):230–238.
6. Michaelson JS, Chen LL, Bush D, et al.Improved web-based calculators for predicting breast carcinoma outcomes. *Breast Cancer Res Treat*. 2011;128(3):827–835.
7. Dellapasqua S, Bagnardi V, Regan MM, et al.A risk score based on histopathological features predicts higher risk of distant recurrence in premenopausal patients with lymph node-negative endocrine-responsive breast cancer. *Breast*. 2012;21(5):621–628.
8. Mazouni C, Spyratos F, Romain S, et al.A nomogram to predict individual prognosis in node-negative breast carcinoma. *Eur J Cancer*. 2012;48(16):2954–2961.
9. Campbell HE, Gray AM, Harris AL, et al.Estimation and external validation of a new prognostic model for predicting recurrence-free survival for early breast cancer patients in the UK. *Br J Cancer*. 2010;103(6):776–786.
10. Paik S, Shak S, Tang G, et al.A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826.
11. van de Vijver MJ, He YD, van 't Veer LJ, et al.A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999–2009.
12. Sotiriou C, Wirapati P, Loi S, et al.Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*. 2006;98(4):262–272.
13. Tutt A, Wang A, Rowland C, et al.Risk estimation of distant metastasis in node-negative, estrogen receptor-positive breast cancer patients using an RT-PCR based prognostic expression signature. *BMC Cancer*. 2008;8.
14. Ma XJ, Salunga R, Dahiya S, et al.A five-gene molecular grade index and HOXB13: IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res*. 2008;14(9):2601–2608.
15. Bartlett JMS, Thomas J, Ross DT, et al.Mammostrat (R) as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res*. 2010;12(4):R47.
16. Parker JS, Mullins M, Cheang MC, et al.Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
17. Harris L, Ismaila N, McShane L, et al.Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol*. 2016;34(10):1134–1150.
18. Hajage D, de Rycke Y, Bollet M, et al.External validation of Adjuvant! Online breast cancer prognosis tool. Prioritising recommendations for improvement. *PLoS One*. 2011;6(11):e27446.
19. Wishart GC, Bajdik CD, Azzato EM, et al.A population-based validation of the prognostic model PREDICT for early breast cancer. *Eur J Surg Oncol*. 2011;37(5):411–417.

ARTICLE

20. Bhoo-Pathy N, Yip CH, Hartman M, et al.Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. *Eur J Cancer*. 2012;48(7): 982–989.

21. Weeks J.Outcomes assessment in the NCCN: 1998 update. National Comprehensive Cancer Network. *Oncology (Williston Park)*. 1999;13(5A):69–71.

22. Christian CK, Niland J, Edge SB, et al.A multi-institutional analysis of the socioeconomic determinants of breast reconstruction: a study of the National Comprehensive Cancer Network. *Ann Surg*. 2006;243(2):241–249.

23. Therneau TM, Grambsh PM.*Modeling Survival Data: Extending the Cox model, Statistics for Biology and Health*. New York: Springer-Verlag; 2000.

24. Iqbal J, Ginsburg O, Rochon PA, et al.Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *JAMA*. 2015;313(2):165–173.

25. Whitman S, Ansell D, Orsi J, et al.The racial disparity in breast cancer mortality. *J Community Health*. 2011;36(4):588–596.

26. Menashe I, Anderson WF, Jatoi I, et al.Underlying causes of the black-white racial disparity in breast cancer mortality: A population-based analysis. *J Natl Cancer Inst*. 2009;101(14):993–1000.

27. Chen L, Li CI.Racial disparities in breast cancer diagnosis and treatment by hormone receptor and HER2 Status. *Cancer Epidemiol Biomarkers Prev*. 2015; 24(11):1666–1672.

28. Janz NK, Mujahid MS, Hawley ST, et al.Racial/ethnic differences in quality of life after diagnosis of breast cancer. *J Cancer Surviv*. 2009;3(4):212–222.

29. Janz NK, Mujahid MS, Hawley ST, et al.Racial/ethnic differences in quality of life and fear of recurrence after diagnosis of breast cancer. *J Clin Oncol*. 2008; 26(15):9526.

30. Murff HJ, Byrne D, Haas JS, et al.Race and family history assessment for breast cancer. *J Gen Int Med*. 2005;20(1):75–80.

31. Ali AM, Schmidt MK, Bolla MK, et al.Alcohol consumption and survival after a breast cancer diagnosis: A literature-based meta-analysis and collaborative analysis of data for 29,239 cases. *Cancer Epidemiol Biomarkers Prev*. 2014;23(6): 934–945.

32. Berube S, Lemieux J, Moore L, et al.Smoking at time of diagnosis and breast cancer-specific survival: New findings and systematic review with meta-analysis. *Breast Cancer Res*. 2014;16(2):R42.

33. Simonsson M, Markkula A, Bendahl PO, et al.Pre- and postoperative alcohol consumption in breast cancer patients: Impact on early events. *Springerplus*. 2014;3:261.

34. Bao PP, Zhao GM, Shu XO, et al.Modifiable lifestyle factors and triple-negative breast cancer survival: A population-based prospective study. *Epidemiology*. 2015;26(6):909–916.

35. Luoma ML, Hakamies-Blomqvist L, Sjostrom J, et al.Prognostic value of quality of life scores for time to progression (TTP) and overall survival time (OS) in advanced breast cancer. *Eur J Cancer*. 2003;39(10):1370–1376.

36. Coates A, Gebski V, Signorini D, et al.Prognostic value of quality-of-life scores during chemotherapy for advanced breast-cancer. *J Clin Oncol*. 1992;10(12): 1833–1838.

37. Efficace F, Biganzoli L, Piccart M, et al.Baseline health-related quality-of-life data as prognostic factors in a phase III multicentre study of women with metastatic breast cancer. *Eur J Cancer*. 2004;40(7):1021–1030.

38. Gupta D, Granick J, Grutsch JF, et al.The prognostic association of health-related quality of life scores with survival in breast cancer. *Support Care Cancer*. 2007;15(4):387–393.

39. Coates AS, Hurny C, Peterson HF, et al.Quality-of-life scores predict outcome in metastatic but not early breast cancer. International Breast Cancer Study Group. *J Clin Oncol*. 2000;18(22):3768–3774.

40. Epplein M, Zheng Y, Zheng W, et al.Quality of life after breast cancer diagnosis and survival. *J Clin Oncol*. 2011;29(4):406–412.

41. De Aguiar SS, Bergmann A, Mattos IE.Quality of life as a predictor of overall survival after breast cancer treatment. *Qual Life Res*. 2014;23(2):627–637.

42. Efficace F, Therasse P, Piccart MJ, et al.Health-related quality of life parameters as prognostic factors in a nonmetastatic breast cancer population: An international multicenter study. *J Clin Oncol*. 2004;22(16):3381–3388.

43. Cook NR.Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928–935.

ARTICLE